Accelerated least squares estimation for systems of ordinary differential equations

Itai Dattner Shota Gugushvili

March 30, 2015

Abstract

We study the problem of parameter estimation for a system of ordinary differential equations based on noisy observations on the solution of the system. A classical estimation approach to this problem is the least squares method. Owing to a highly nonlinear character of the least squares criterion function and the need to employ repetitive numerical integration of the system, the latter method becomes computationally intense for most realistic systems. We propose the accelerated (ACCEL) least squares method, which starts from a preliminary $\sqrt{n}$-consistent estimator of the parameter of interest and next through a Newton-Raphson type step turns it into an asymptotically equivalent estimator to the least squares estimator. Additional computational burden of this step is negligible. We demonstrate excellent practical performance of the ACCEL least squares approach via simulations and real data examples. The method enables the researcher to obtain both point and interval estimates.

March 30, 2015

1 Introduction

Systems of ordinary differential equations (ODEs) are commonly used for the mathematical modeling of the rate of change of dynamic processes (e.g., in mathematical biology, see Edelstein-Keshet (2005); in the theory of chemical reaction networks, see Feinberg (1979) and Sontag (2001); and in biochemistry, see Voit (2000)). The systems we have in mind take the form

$$\begin{cases} x'(t) = F(x(t),\theta), \quad t \in [0,1], \\ x(0) = \xi \end{cases} \tag{1}$$

where $x(t) = (x_1(t),\dots,x_d(t))^T$ is a d-dimensional state variable, $\theta = (\theta_1,\dots,\theta_p)^T$ denotes a p-dimensional parameter, while the column d-vector $x(0) = \xi$ defines the initial condition. We define $\eta := (\xi,\theta)$ and denote the solution to (1) corresponding to the parameter η by

$$x(\eta,t) := (x_1(\eta,t),\dots,x_d(\eta,t))^T.$$

Knowledge regarding the system parameters ξ and θ is of vital importance for the study of a process that (1) models. Indeed, these parameters affect the qualitative properties of the system, and their knowledge allows one to predict the system behaviour. However, in practice the parameter θ, and possibly also ξ, are unknown to the researcher. Typically they cannot be measured directly, but have to be inferred from noisy measurements of the process under study. Let $\eta_0 = (\xi_0, \theta_0)$ be the 'true' parameter value that governs the underlying process. The common statistical model considered for the noisy measurements of the process at time instances $t_1, \ldots, t_n$ (deterministic or random; not necessarily equally spaced), is the additive measurement error model,

$$Y_{ij} = x_i(\eta_0, t_j) + \varepsilon_{ij}, \quad i = 1, \ldots, d, j = 1, \ldots, n, \tag{2}$$

where the random variables ε_{ij} are independent measurement errors (not necessarily Gaussian). Based on observation pairs (t_j, Y_{ij}), $i = 1, \ldots, d$, $j = 1, \ldots, n$, the goal is to estimate the parameter η_0.

A classical approach to parameter estimation for ordinary differential equations is the nonlinear least squares (NLS) method. Its use is based on the observation that the problem at hand in its essence is a nonlinear regression problem, where the regression function $x(\eta, \cdot)$ is defined implicitly as the solution to (1). The least squares estimator $\widetilde{\eta}_n = (\widetilde{\xi}_n, \widetilde{\theta}_n)$ of η_0 is defined as a minimiser of the least squares criterion function $R_n(\cdot)$,

$$\begin{aligned} \widetilde{\eta}_n \quad &= \quad (\widetilde{\xi}_n, \widetilde{\theta}_n) = \operatorname{argmin}_\eta \sum_{i=1}^{d} \sum_{j=1}^{n} (Y_{ij} - x_i(\eta, t_j))^2 \\ &=: \operatorname{argmin}_\eta R_n(\eta). \end{aligned} \tag{3}$$

The strongest justification for the use of the least squares estimator lies in its attractive asymptotic properties; see e.g., Jennrich (1969) and Wu (1981). In most practical applications the solution $x(\eta, \cdot)$ to (1) is nonlinear in the parameter η, and therefore some iterative procedure has to be used to compute the nonlinear least squares estimator. Such procedures require an initial guess for a minimiser $\widetilde{\eta}_n$ and then proceed by constructing successive approximations to the least squares estimator (in a direction guided by the gradient of the criterion function, when a gradient based optimisation method, e.g. the Levenberg-Marquardt method, is used). However, the noisy and nonlinear character of the optimization problem may lead for the procedure to end up in a local minimum of the least squares criterion function, especially when good initial guesses of the parameter values are not available. Furthermore, in most applications the system (1) is nonlinear and does not have a closed form solution. In that case at every step of the iterative procedure one has to numerically integrate (1) (as well as the system of the associated sensitivity equations in order to compute the gradient of the criterion function, in case a gradient-based optimisation method is used). Since the number of iterations made until convergence of the algorithm can be ascertained is usually large, in most cases this leads to a computational bottleneck. This is the case especially in mathematical biology and biochemistry, where a highly nonlinear character of dependence of the solution $x(\eta, \cdot)$ on the parameter η leads to 'stiff' integration problems. For a penetrating discussion of these point see e.g. Ramsay et al. (2007) and Voit and Almeida (2004).

A Bayesian approach to parameter estimation for ordinary differential equations is a viable alternative to the NLS method; see, e.g., Gelman et al. (1996), Girolami (2008), Girolami and Calderhead (2011). However, it will not be discussed in the present work. We only remark that like the NLS method, its application also requires a considerable computational effort.

In order to bypass computational problems associated with computation of the least squares estimator, Bellman and Roth (1971) and Varah (1982) suggested to estimate the solution $x(\eta_0,\cdot)$ using nonparametric techniques (e.g., splines). This approach was studied rigorously in Gugushvili and Klaassen (2012) (other relevant references are e.g. Brunel (2008), Vujačić et al. (2014) and Dattner and Klaassen (2015)) and works as follows: the observations are first smoothed, which results in an estimator $\widehat{x}_n(\cdot)$ for the solution $x(\eta_0,\cdot)$ of the system, and by differentiation, in an estimator $\widehat{x}_n'(\cdot)$ for $x'(\eta_0,\cdot)$. Then the estimator for θ_0 is the minimizer $\widehat{\theta}_n$ over θ of the smooth function

$$\int_0^1 \| \widehat{x}_n'(t) - F(\widehat{x}_n(t);\theta) \|^2 \, w(t)\,\mathrm{d}t, \tag{4}$$

where w is an appropriate weight function, and $\|\cdot\|$ denotes the standard Euclidean norm. Hence, this approach bypasses the need to integrate the system numerically, and as a result the parameter estimates can be computed extremely quickly, especially when F in (1) is linear in θ. Under regularity conditions Gugushvili and Klaassen (2012) show that this *smooth and match estimator* (SME) $\widehat{\theta}_n$ has the $\sqrt{n}$-rate of convergence to θ. By the general statistical theory, the $\sqrt{n}$-rate of convergence is in fact the best rate one can expect in the present context. This result thus puts the smooth and match method on a solid theoretical ground.

Note that execution of this method does not require the knowledge of the initial values. However, it cannot be used to estimate initial values. If estimation of initial values is of interest, then once the estimator $\widehat{\theta}_n$ is at hand, one may obtain an estimator $\widehat{\xi}_n$ by minimizing with respect to ξ the criterion

$$\int_0^1 \| \widehat{x}_n(t) - \xi - \int_0^t F(\widehat{x}_n(s);\widehat{\theta}_n)\mathrm{d}s \|^2 \,\mathrm{d}t.$$

Notice that this is a linear least squares optimization problem and hence is easy to execute.

The SME described above is criticized for not being *statistically efficient*. In informal terms this means that the resulting estimators do not squeeze as much information out of the data as the least squares estimator does (in more formal terms, their asymptotic variance is larger than that of the least squares estimator). Hence, sometimes it is suggested (see, e.g., Swartz and Bremermann (1975) for an early reference) to use this method only for generating preliminary estimates that should be used later as an initial guess for more accurate methods, that, unfortunately, are not computationally efficient. In this paper we present a new algorithm, ACCEL (accelerated) least squares, that enjoys both the computational efficiency of the SME approach and the statistical efficiency of the least squares estimators. The method is studied theoretically in an unpublished report by Gugushvili and Klaassen, while here we will concentrate on numerical aspects.

The rest of the paper is organized as follows: in Section 2 we describe the ACCEL least squares algorithm. Section 3 presents a simulation study illustrating the performance of our method, while Section 4 contains numerical results based on real data examples. Section 5 summarizes our contribution and outlines potential future research directions.

2 Accelerated least squares estimator

When one adpots an asymptotic point of view on statistics, all the estimators with the same asymptotic variance can be considered as equivalent. We now demonstrate how once some $\sqrt{n}$-consistent estimator $\widehat{\eta}_n$ of the parameter η is available (e.g. the SME), one can obtain an asymptotically equivalent estimator to the least squares estimator in just one extra easy step. The method is referred to as the one-step method in the statistical literature, see e.g. Section 5.7 in van der Vaart (1998).

Consider the estimating equations

$$\Psi_n(\eta) = \sum_{j=1}^{n} \psi_\eta(t_j, Y_j) = 0, \tag{5}$$

where

$$\psi_\eta(t,y) = (y - x(\eta,t))(x'_\eta(\eta,t))^T,$$

with $x'_\eta(\eta,t)$ denoting the derivative of $x(\eta,t)$ with respect to η. Up to omission of a constant multiplier -2, these estimating equations are obtained by differentiating the least squares criterion function $R_n(\eta)$ from (3) with respect to η. A value of $\eta = (\xi, \theta)$ that solves (5) is called a Z-estimator. In general there can be multiple solutions to (5), one of them being the least squares estimator. Since η is a $p+d$-dimensional vector, we will need $p+d$ estimating equations, so that $\Psi_n(\eta) = 0$ stands for $p+d$ equations.

The accelerated (ACCEL) least squares estimator $\overline{\eta}_n$ of η_0 is defined as a solution in η of the equation

$$\Psi_n(\widehat{\eta}_n) + \frac{d}{d\eta}\Psi_n(\widehat{\eta}_n)(\eta - \widehat{\eta}_n) = 0.$$

This corresponds to replacing $\Psi_n(\eta)$ by its tangent line at $\widehat{\eta}_n$, and $\overline{\eta}_n$ is then the 'η-intercept' of the tangent line. If $\frac{d}{d\eta}\Psi_n(\widehat{\eta}_n)$ is invertible, the estimator $\overline{\eta}_n$ can be expressed as

$$\overline{\eta}_n = \widehat{\eta}_n - \left(\frac{d}{d\eta}\Psi_n(\widehat{\eta}_n)\right)^{-1}\Psi_n(\widehat{\eta}_n).$$

If $\widehat{\eta}_n$ is 'close' to the solution of the estimating equations (5), the ACCEL estimator $\overline{\eta}_n$ will also be 'close' to it, and hopefully, even 'closer' than $\widehat{\eta}_n$. In the numerical analysis, when looking for a zero of an equation, such a procedure is iterated several times, each time taking the previously found value as a new starting point, and is known under the name of Newton's method, see e.g., Section 2.3 in Burden and Faires (2001). On the other hand, the asymptotic theory of statistics tells us that given certain regularity assumptions on the statistical model, the estimator $\overline{\eta}_n$ is just as accurate asymptotically as a Z-estimator obtained from solving $\Psi_n(\eta) = 0$. Thus from the asymptotic point of

view there is no difference between the least squares estimator and the estimator $\overline{\eta}_n$: both are equally accurate. At the same time, the extra computational cost needed to evaluate the ACCEL estimator $\overline{\eta}_n$ is negligible. Indeed, the only essential additional thing that has to be done for its computation in comparison to computation of the SME is the evaluation of $\Psi_n(\widehat{\eta}_n)$ and the derivative matrix $\frac{d}{d\eta}\Psi_n(\widehat{\eta}_n)$. This reduces to requiring just one numerical integration of the sensitivity and variational equations associated with (1), as we will now illustrate.

In what follows, it is helpful to think of F in (1) as a function of η rather than only θ. Thus, we write the right-hand side F of (1) as $F(x(\eta,t),\eta)$. Differentiating both sides of (1) with respect to η and interchanging the order of a t-derivative with an η-derivative, we get

$$\begin{cases} \frac{d}{dt}\frac{\partial}{\partial\eta}x(\eta,t) = F'_x(x(\eta,t),\eta)\frac{\partial}{\partial\eta}x(\eta,t) + F'_\eta(x(\eta,t),\eta), \\ \frac{\partial}{\partial\eta}x(\eta,0) = (1,0)^T. \end{cases} \tag{6}$$

This system is a matrix differential equations and is usually referred to in the literature as a system of sensitivity equations. By replacing η with $\widehat{\eta}_n$ we arrive at the system

$$\begin{cases} \frac{d}{dt}s(t) = F'_x(x(\widehat{\eta}_n,t),\widehat{\eta}_n)s(t) + F'_\eta(x(\widehat{\eta}_n,t),\widehat{\eta}_n), \\ s(0) = (1,0)^T, \end{cases} \tag{7}$$

where we have defined $s(t) := \frac{d}{d\eta}x(\widehat{\eta}_n,t)$. Observe that $x(\widehat{\eta}_n,\cdot)$ is a known function, because it can be found by integrating (1) for parameter values $\widehat{\xi}_n$ and $\widehat{\theta}_n$. Consequently, the system of sensitivity equations is a linear system with time-dependent coefficients, and hence is relatively straightforward to integrate.

By differentiating (6) one more time with respect to η and replacing η with $\widehat{\eta}_n$ we arrive at the following set of variational equations:

$$\begin{cases} \frac{d}{dt}z(t) = F''_{\eta\eta}(x(\widehat{\eta}_n,t),\widehat{\eta}_n)) + F''_{\eta x}(x(\widehat{\eta}_n,t),\widehat{\eta}_n)s(t) \\ +\left\{F''_{x\eta}(x(\widehat{\eta}_n,t),\widehat{\eta}_n) + F''_{xx}(x(\widehat{\eta}_n,t),\widehat{\eta}_n)s(t)\right\}s(t) \\ +F'_x(x(\widehat{\eta}_n,t),\widehat{\eta}_n)z(t), \\ z(0) = 0, \end{cases} \tag{8}$$

where we have set $z(t) := \frac{\partial^2}{\partial\eta^2}x(\eta,t)$. For each z_i, $i = 1,\ldots,d$, the system (8) is a matrix differential equation and again is a linear system with time-varying coefficients. Here also we can treat x and s as known, for they can be obtained through numerical integration of (1) and (7). The process of obtaining variational equations can be made automatic through a software implementation, and we provide such a Matlab code.

Integration of (1), (7) and (8) for the parameter value $\widehat{\eta}_n$ will allow us to compute $\Psi_n(\widehat{\eta}_n)$ and $\frac{d}{d\eta}\Psi_n(\widehat{\eta}_n)$, and consequently, the ACCEL estimator $\overline{\eta}_n$. Note that numerical integration of the variational equations (or at least the sensitivity equations) is usually required when computing the least squares estimator via gradient-based optimization methods (unless the gradient is available analytically) or using the Bayesian approach (see Girolami and Calderhead (2011)). However, we need to do this only *once*.

Remark 1 *A seemingly more general nonautonomous system than the autonomous system* (1)*,*

$$\begin{cases} \widetilde{x}'(t) = F(\widetilde{x}(t), t; \theta), \ t \in [0,1], \\ \widetilde{x}(0) = \widetilde{\xi}, \end{cases}$$

may and will be reduced to (1) *by a simple substitution* $x(t) = (\widetilde{x}^T(t), t)^T, \ t \in [0,1]$, *and* $\xi = (\widetilde{\xi}^T, 0)^T$.

3 Simulation study

3.1 Linear ODE

We start with illustrating the performance of the ACCEL estimator when used to estimate the parameter and initial value of a one-dimensional linear ordinary differential equation

$$\begin{cases} x'(t) = \theta_0 x(t), \ t \in [0,T], \\ x(0) = \xi_0. \end{cases} \tag{9}$$

This is a toy example, but it allows us to explore the practical performance of the ACCEL method in great detail and to compare it to the theoretically expected results. Advanced examples will be considered later on.

Given observations Y_{ij}'s, the ACCEL method requires first to have at hand a $\sqrt{n}$-rate estimator of θ_0 and ξ_0. As mentioned in the previous sections, the SME provides us with such an estimator. However, this method is based on estimating the derivative x', which is hard to do accurately in practice. In the case where the symbol of the system of ODEs is linear in functions of the parameter θ, as is the case in (9), one can avoid the estimation of derivatives and use an integral SME. Indeed, in such cases one can use some version of the so called 'integral approach' (Himmelblau et al. (1967)) as was studied in Dattner and Klaassen (2015). For systems linear in the parameter θ as in (9) the idea works as follows.

Note that in (9) $F(x(t);\theta) = g(x(t))\theta$ holds, where the measurable function $g: \mathbb{R}^d \to \mathbb{R}^{d\times p}$ maps the d-dimensional column vector x into a $d \times p$ matrix. Let $\widehat{x}_n(\cdot)$ be an estimator of $x(\eta_0, \cdot)$, and denote $\widehat{G}_n(t) = \int_0^t g(\widehat{x}_n(s), s)\mathrm{d}s$, $\widehat{A}_n = \int_0^T \widehat{G}_n(t)\mathrm{d}t$, $\widehat{B}_n = \int_0^T \widehat{G}_n^T(t)\widehat{G}_n(t)\mathrm{d}t$, and I_d be the $d \times d$ identity matrix. Then Dattner and Klaassen (2015) show that the direct estimators

$$\widehat{\xi}_n = \left(I_d - \widehat{A}_n\widehat{B}_n^{-1}\widehat{A}_n^T\right)^{-1} \int_0^T \left(I_d - \widehat{A}_n\widehat{B}_n^{-1}\widehat{G}_n^T(t)\right)\widehat{x}_n(t)\,\mathrm{d}t, \tag{10}$$

$$\widehat{\theta}_n = \widehat{B}_n^{-1} \int_0^T \widehat{G}_n^T(t)\left(\widehat{x}_n(t) - \widehat{\xi}_n\right)\mathrm{d}t, \tag{11}$$

are $\sqrt{n}$-consistent. In case the initial value ξ_0 is known, (11) may be used with $\widehat{\xi}_n$ replaced by ξ_0. Besides the required statistical properties, the extensive simulation study presented in the aforementioned paper suggests that this approach is much more accurate in finite samples comparing to the derivative SME resulting from minimizing

(4). Thus, we use the integral SME (10)-(11) whenever applicable, and the derivative SME (4) otherwise.

We choose to estimate the solution x using local polynomial estimators, which are consistent and 'automatically' correct for the boundaries. Under the assumption that x are C^{α}-functions for some real $\alpha \geq 1$, we will approximate them by polynomials of degree $\ell = \lfloor \alpha \rfloor$ as follows (Tsybakov 2009, Section 1.6). Let

$$\begin{aligned} U(u) &= \Big(1, u, u^2/(2!), \ldots, u^{\ell}/(\ell!)\Big)^T, \quad u \in \mathbb{R}, \\ \nu(t) &= \Big(x(t), x'(t)b, x''(t)b^2, \ldots, x^{(\ell)}(t)b^{\ell}\Big), \quad t \in \mathbb{R}, \end{aligned}$$

where $b = b_n > 0$ is a bandwidth, the $(\ell+1)$-vector $U(u)$ is a column vector, and $\nu(t)$ is a $d \times (\ell+1)$-matrix. Let $K(\cdot)$ be some appropriate kernel function and define

$$\begin{aligned} \widehat{\nu}_n(t) &= \arg\min_{\nu \in \mathbb{R}^{d\times(\ell+1)}} \sum_{i=1}^{n} \Big\{ Y(t_i) - \nu U\Big(\frac{t_i - t}{b}\Big)\Big\}^T \\ &\quad \times \Big\{ Y(t_i) - \nu U\Big(\frac{t_i - t}{b}\Big)\Big\} K\Big(\frac{t_i - t}{b}\Big). \end{aligned}$$

The local polynomial estimator of order ℓ of $x(t)$ is the first column of the $d \times (\ell+1)$-matrix $\widehat{\nu}_n(t)$, i.e., $\widehat{x}_n(t) = \widehat{\nu}_n(t) U(0)$.

In the sequel we apply the estimation procedure described above to a set of bandwidths $B := \{b_{\min}, \ldots, b_{\max}\}$, and for a given $b \in B$ we denote the resulting ACCEL parameter estimator by $\overline{\eta}_{n,b}$. We then choose

$$\overline{\eta}_n = \overline{\eta}_{n,\bar{b}}, \tag{12}$$

where

$$\bar{b} = \arg\min_{b \in B} \sum_{i=1}^{d} \sum_{j=1}^{n} (Y_{ij} - x_i(\overline{\eta}_{n,b}, t_j))^2.$$

Here $x_i(\overline{\eta}_{n,b}, \cdot)$ is the solution of (1) using the ACCEL estimator $\overline{\eta}_{n,b}$.

The theoretical analysis of Dattner and Klaassen (2015) shows that in order to have the $\sqrt{n}$-rate for the parameter estimators, once should take a bandwidth $b = O(n^{-1/3})$. Thus, we set $B = n^{-1/3} \times (c_1, \ldots, c_N)$, N a positive integer, and the c_j's depend on the grid of points on which we evaluate the kernel estimator. Last, we use local estimators polynomials of order 1, with $K(t) = 3/4(1-t^2)\mathbf{1}\{|t| \leq 1\}$ (cf., Dattner and Klaassen (2015)), where $\mathbf{1}\{\cdot\}$ stands for the indicator function. Other kernels are also possible.

The solution of the initial value problem (9) is $x(t) = x_0 \exp(\theta_0 t)$. We generate (pseudo) random observations from the model

$$Y_j = \xi_0 \exp(\theta_0 t_j) + \varepsilon_j,$$

where $t_j \in \{0(0.1)10\}$ ($n = 101$), and $\varepsilon_j \sim N(0, 0.05^2)$, $j = 1 \ldots, n$. We consider

$$\theta_0 \in \{-1, -0.8, -0.6, -0.4, -0.2, 0.1, 0.3, 0.5, 0.7, 0.9\}$$

and $\xi_0 \in \{0.5, 1\}$. For each pair (ξ_0, θ_0) we run a Monte Carlo study of 500 samples of $Y_1 \ldots, Y_{101}$, where in each sample we apply the ACCEL least squares method as described above, and the nonlinear least squares. We note that we observed some setups, where the ACCEL estimator may perform differently, depending whether initial values are estimated or not. Thus, in this example we used scaling of the time t, such that $t \in [0,1]$, which stabilized the procedure. However, scaling was not needed in all other examples we consider below. This simulation study enables us to estimate the asymptotic variance of the least squares and the ACCEL least squares methods. We then compare the results to the true asymptotic variance. The true and estimated asymptotic variances can be obtained for each set of parameters and initial values by inverting the Fisher information matrix. Assuming the observation model (2), where we have homoscedastic errors, and assuming that the observed time points $t_1, \ldots, t_n$ are uniformly distributed over $[0, T]$, the Fisher information matrix is given by the $(d+p) \times (d+p)$ matrix

$$I(\eta) = \frac{1}{\sigma^2} \sum_{i=1}^{d} \frac{1}{T} \int_0^T \left(\frac{d}{d\eta} x_i(\eta, t) \right) \left(\frac{d}{d\eta} x_i(\eta, t)^T \right) \mathrm{d}t, \tag{13}$$

where we can obtain $s(t) := \frac{d}{d\eta} x(\eta, t)$ as in (7). However, the Fisher information matrix depends on the true values of the parameters, initial values, and σ^2, which are not known in practice. So in general, we will construct a fully data driven confidence interval by estimating the Fisher information matrix. We estimate σ^2 by

$$\widehat{\sigma}^2 = \frac{1}{d(n-1)} \sum_{i=1}^{d} \sum_{j=1}^{n} (Y_{ij} - x_i(\overline{\eta}_n, t_j))^2,$$

where $x(\overline{\eta}_n, \cdot)$ stands for the solution of the system (9) using the estimated parameters and initial values using the ACCEL method. Then an estimate for the asymptotic variance of the parameter η_j is given by $I^{-1}_{jj}(\overline{\eta}_n)/n$,where $I^{-1}_{jj}(\overline{\eta}_n)$ stands for the jth diagonal element of the inverse Fisher information matrix evaluated in point $\overline{\eta}_n$. When $s(\cdot)$ has no closed form, the integral in (13) will be evaluated using numerical integration (in the subsequent examples we use the trapezoidal rule).

A direct computation gives that in model (9) the asymptotic variance of $\overline{\xi}_n$ depends on θ, but is independent of the values of ξ itself. In Figure 1 we see the estimated variance of the ACCEL estimators (plus signs) and that of the NLS (circles), for estimating ξ_0 based on 500 simulation runs. The estimates are superimposed on the theoretical asymptotic variance (dashed line). The top plot is for $\xi_0 = 0.5$ and the bottom one is for $\xi_0 = 1$. As the theory suggests, independently of the values of ξ, the true asymptotic variance is the same. Note that in this specific numerical example the estimated variances of the ACCEL and NLS estimators are the same. This is not surprising, since in order to apply the NLS we used as the initial point in the parameter space the SME (resulted from using a kernel with the largest bandwidth; this choice was arbitrary). The estimated variances agree with the asymptotic one. We note that the grid of θ_0 does not include 0, where the asymptotic variance equals zero.

In Figure 2 we see similar plots corresponding to estimating the asymptotic variances of $\overline{\theta}_n$. Here the variance has different order, depending on the value of ξ_0. Again,

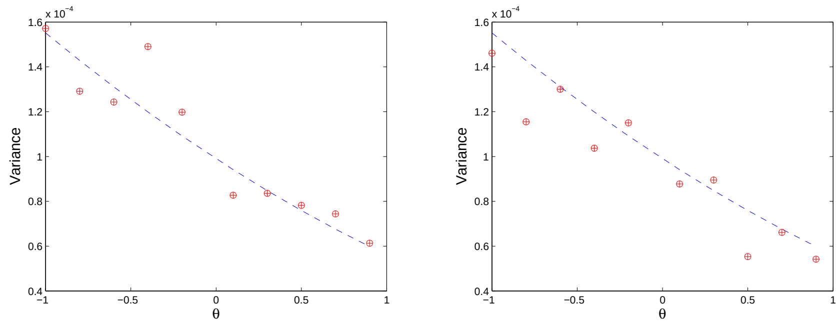

Figure 1: The estimated variance of the ACCEL (plus signs) and NLS (circles) estimators $\overline{\xi}_n$ and $\widetilde{\xi}_n$, respectively, based on 500 simulations with $n = 101$ and $\varepsilon_j \sim N(0, 0.05^2)$, $j = 1\ldots,n$. The estimates are superimposed on the theoretical asymptotic variance (dashed line). The upper plot is for $\xi_0 = 0.5$ and the lower one is for $\xi_0 = 1$.

the estimated variances of the ACCEL (plus signs) and NLS (circles) estimators are the same, and both agree with the asymptotic one (dashed line). Similar plots were obtained when considering other values for σ^2, and therefore we do not present them here.

The discussion above implies that we can provide confidence intervals together with the point estimates. An approximate $1-\alpha$ level confidence interval for η_j is given by

$$[\overline{\eta}_{j,n} - z_{1-\alpha} I_{jj}^{-1}(\overline{\eta}_{j,n})/n, \overline{\eta}_n - z_{1-\alpha} I_{jj}^{-1}(\overline{\eta}_{j,n})/n], \tag{14}$$

where $z_{1-\alpha}$ is the $1-\alpha$ quantile of the standard normal distribution. In Table 1 we present the coverage of this confidence interval based on a Monte Carlo study with 500 simulations for different experimental setups. The results should be compared to the nominal coverage of 95%. We consider 4 setups denoted by A, B, C, D according to $(\xi_0 = 1/2, \theta_0 = -1), (\xi_0 = 1/2, \theta_0 = 1), (\xi_0 = 1, \theta_0 = -1), (\xi_0 = 1, \theta_0 = -1)$, respectively. Each scenario is tested for $n = 21$, and $n = 51$. Table 1 presents the point and interval estimates for the parameters of each scenario. We see that the coverage of the data driven confidence intervals is satisfying across the different experimental scenarios.

We close this example by the following illustration. In Table 2 we present for both ACCEL and NLS methods, the square root of the average of the estimates of the asymptotic variance, over the 500 simulations (denoted by 'ASYM'). Next to that we present standard errors of the point estimates as calculated based on the 500 simulations (denoted by 'STE'). The results agree with each other.

3.2 Lotka-Volterra system

The Lotka-Volterra system of ODEs (Edelstein-Keshet (2005)) is a population dynamics model that describes evolution over time of the populations of two species, predators

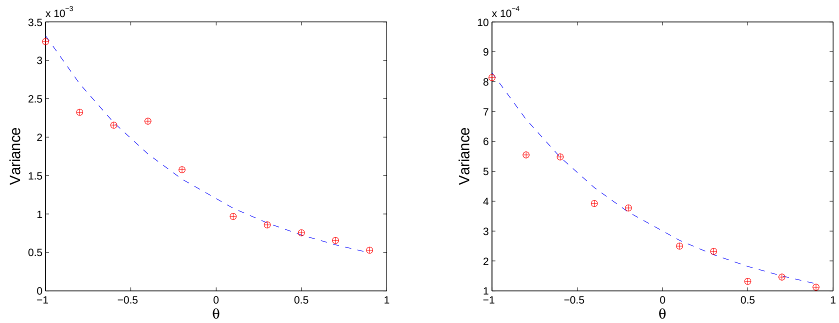

Figure 2: The estimated variance of the ACCEL (plus signs) and NLS (circles) estimators $\bar{\theta}_n$ and $\widetilde{\theta}_n$, respectively, based on 500 simulations with $n = 101$ and $\varepsilon_j \sim N(0, 0.05^2)$, $j = 1\dots,n$. The estimates are superimposed on the theoretical asymptotic variance (dashed line). The upper plot is for $\xi_0 = 0.5$ and the lower one is for $\xi_0 = 1$.

Table 1: Means of point estimates and actual coverage of interval estimates for the parameters of model (9) according to 4 different experimental setups. The results are based on 500 simulation runs. The observations are generated according to $Y_j = \xi_0 \exp(\theta_0 t_j) + \varepsilon_j$, where $t_j \in \{0(0.5)10\}$ ($n = 21$), or $t_j \in \{0(0.2)10\}$ ($n = 51$) and $\varepsilon_j \sim N(0, 0.05^2)$, $j = 1\dots,n$. The point estimates are given by (12); the interval estimates are defined in (14)

				ACCEL		NLS	
Setup				Mean	Coverage	Mean	Coverage
n=21	A	ξ_0	0.500	0.501	0.942	0.501	0.942
		θ_0	-1.000	-1.002	0.946	-1.002	0.946
	B	ξ_0	0.500	0.500	0.928	0.500	0.928
		θ_0	1.000	1.000	0.938	1.000	0.938
	C	ξ_0	1.000	0.999	0.932	0.999	0.932
		θ_0	-1.000	-0.997	0.940	-0.997	0.940
	D	ξ_0	1.000	1.000	0.944	1.000	0.944
		θ_0	1.000	1.000	0.948	1.000	0.948
n=51	A	ξ_0	0.500	0.500	0.944	0.500	0.944
		θ_0	-1.000	-0.998	0.944	-0.998	0.944
	B	ξ_0	0.500	0.500	0.946	0.500	0.946
		θ_0	1.000	0.999	0.958	0.999	0.958
	C	ξ_0	1.000	0.999	0.932	0.999	0.932
		θ_0	-1.000	-0.999	0.938	-1.000	0.938
	D	ξ_0	1.000	1.000	0.948	1.000	0.948
		θ_0	1.000	1.001	0.952	1.001	0.952

Table 2: Standard errors of the point estimates as calculated based on the 500 simulations (denoted by 'STE'). Square root of the average of the estimates of the asymptotic variance, over the 500 simulations (denoted by 'ASYM').

				ACCEL		NLS	
Setup				STE	ASYM	STE	ASYM
n=21	A	ξ_0	0.500	0.027	0.031	0.027	0.031
		θ_0	-1.000	0.121	0.140	0.121	0.140
	B	ξ_0	0.500	0.016	0.012	0.016	0.012
		θ_0	1.000	0.045	0.034	0.045	0.034
	C	ξ_0	1.000	0.027	0.034	0.027	0.034
		θ_0	-1.000	0.061	0.079	0.061	0.079
	D	ξ_0	1.000	0.016	0.016	0.016	0.016
		θ_0	1.000	0.022	0.023	0.022	0.023
n=51	A	ξ_0	0.500	0.017	0.017	0.017	0.017
		θ_0	-1.000	0.080	0.076	0.080	0.076
	B	ξ_0	0.500	0.010	0.009	0.010	0.009
		θ_0	1.000	0.028	0.026	0.028	0.026
	C	ξ_0	1.000	0.017	0.019	0.017	0.019
		θ_0	-1.000	0.040	0.043	0.040	0.043
	D	ξ_0	1.000	0.010	0.010	0.010	0.010
		θ_0	1.000	0.015	0.014	0.015	0.014

and their preys. The system takes the form

$$\begin{cases} x_1'(t) = \theta_1 x_1(t) - \theta_2 x_1(t)x_2(t), \\ x_2'(t) = -\theta_3 x_2(t) + \theta_4 x_1(t)x_2(t). \end{cases} \tag{15}$$

Here x_1 represents the size of the prey population and x_2 of the predator population. In Table 3 we see the empirical coverage of the 95% confidence intervals based on a Monte Carlo study consisting of 500 simulation runs for different sample sizes. The experimental setup is as follows: the observed time points are equidistant on $[0,10]$; the errors are normal with zero mean and standard deviation $\sigma = 0.05$; the initial values are $\xi_0 = (1,1/2)^T$, and the parameters are $\theta_0 = (1/2,1/2,1/2,1/2)^T$. The point estimates are given by (12), while the interval estimates are defined in (14). As expected, the coverage is much better when the sample size is larger. The performance of the ACCEL and NLS methods is similar.

Last, for both ACCEL and NLS methods, in Table 4 we present the square root of the average of the estimates of the asymptotic variance, over the 500 simulations (denoted by 'ASYM'). Next to that we present standard errors of the point estimates as calculated based on the 500 simulations (denoted by 'STE'). The results agree with each other.

Table 3: Means of point estimates and actual coverage of interval estimates for the parameters of model (15), where the initial values are $\xi = (1, 1/2)^T$, and the rate parameters are $\theta_0 = (1/2, 1/2, 1/2, 1/2)^T$. The results are based on running 500 simulations. The observed time points are equidistant on $[0, 10]$, and the errors are normal with zero expectation and $\sigma = 0.05$. The ACCEL point estimates are given by (12); the interval estimates are defined in (14).

			ACCEL		NLS	
Setup			Mean	Coverage	Mean	Coverage
n=21	ξ_1	1.000	1.000	0.932	0.999	0.928
	ξ_2	0.500	0.500	0.936	0.500	0.934
	θ_1	0.500	0.502	0.942	0.501	0.942
	θ_2	0.500	0.502	0.932	0.501	0.938
	θ_3	0.500	0.500	0.910	0.501	0.916
	θ_4	0.500	0.500	0.918	0.501	0.922
n=51	ξ_1	1.000	1.000	0.958	1.000	0.966
	ξ_2	0.500	0.500	0.954	0.500	0.948
	θ_1	0.500	0.502	0.964	0.500	0.968
	θ_2	0.500	0.501	0.968	0.500	0.964
	θ_3	0.500	0.500	0.958	0.500	0.958
	θ_4	0.500	0.500	0.952	0.500	0.958

Table 4: Standard errors of the point estimates as calculated based on the 500 simulations (denoted by 'STE'). Square root of the average of the estimates of the asymptotic variance, over the 500 simulations (denoted by 'ASYM').

		ACCEL		NLS	
Setup		STE	ASYM	STE	ASYM
n=21	ξ_1	0.025	0.023	0.025	0.023
	ξ_2	0.020	0.019	0.020	0.019
	θ_1	0.027	0.026	0.027	0.026
	θ_2	0.022	0.021	0.022	0.021
	θ_3	0.022	0.020	0.022	0.020
	θ_4	0.020	0.018	0.020	0.018
n=51	ξ_1	0.015	0.016	0.014	0.016
	ξ_2	0.013	0.013	0.013	0.013
	θ_1	0.016	0.017	0.016	0.017
	θ_2	0.013	0.014	0.013	0.014
	θ_3	0.013	0.013	0.013	0.014
	θ_4	0.012	0.012	0.012	0.012

4 Real data examples

In this section we study several realistic examples, also employing real data. To check applicability of our method, the emphasis will be on examples with small or moderate sample sizes.

4.1 Nitrogene oxide reaction

The system

$$\begin{cases} x'(t) = \theta_1(126.2 - x(t))(91.9 - x(t))^2 - \theta_2(x(t))^2, \\ x(0) = 0 \end{cases} \tag{16}$$

describes the reversible homogeneous gas phase reaction of nitrogene oxide,

$$2\mathrm{NO} + \mathrm{O}_2 \rightleftharpoons 2\mathrm{NO}_2.$$

The forward reaction is the third order, while the reverse reaction is the second order. Time in (16) is measured in minutes and the variable $x(t)$ models pressure fall at time t. For additional chemical background see Bodenstein (1922). Based on the experimental data from Table 39 in Bodenstein (1922), parameters of equation (16) were estimated via different methods in Bellman et al. (1967); Van Domselaar and Hemker (1975), see pp. 18–19; Esposito and Floudas (2000), Section 7.4; Kim and Sheng (2010), Section 3.1; Tjoa and Biegler (1991), Problem 6 on p. 381; and Varah (1982), see pp. 37–38. The results obtained in these papers are summarised in Table 5.[1] We also remark that this problem is one of the six test problems in parameter estimation for ordinary differential equations that were included in Floudas (1999).

Table 5: Parameter estimates for model (16) obtained in the literature.

Paper	Estimate of θ_1	Estimate of θ_2
Bellman et al. (1967)	0.4577×10^{-5}	0.2797×10^{-3}
Van Domselaar and Hemker (1975)	0.45×10^{-5}	0.27×10^{-3}
Esposito and Floudas (2000)	0.4593×10^{-5}	0.28285×10^{-3}
Kim and Sheng (2010)	0.46×10^{-5}	0.28×10^{-3}
Tjoa and Biegler (1991)	0.4604×10^{-5}	0.2847×10^{-3}
Varah (1982)	0.46×10^{-5}	0.27×10^{-3}

Our interest in this example first went in the following direction: we used the realistic estimated parameter values from the literature, generated an artificial set of data from (16) and checked how well the ACCEL estimator performs in this case. We

[1] Note that Varah (1982) gives five different parameter estimates corresponding to different values of the smoothing parameter used in his method. Of these estimates we report only the first pair and refer to Table 4 in Varah (1982) for the remaining ones. Note also that Esposito and Floudas (2000) use two approaches (collocation method and integration method in their terminology) and with the second of them identify another local solution to the problem, namely $\theta_1 = 0.1306 \times 10^{-2}, \theta_2 = 0.90393$ (see Table 11 in Esposito and Floudas (2000)), which we did not report in Table 5.

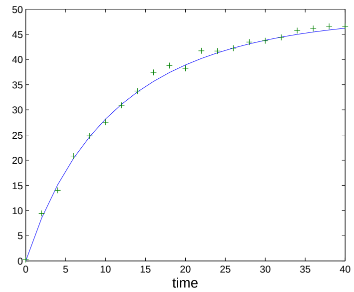

Figure 3: The solution to (16) is given by the solid line, while the observations are indicated by pluses. The parameter is $\eta_0 = (\xi, \theta_1, \theta_2)^T = (0, 0.4577 \times 10^{-5}, 0.2797 \times 10^{-3})^T$, and the observations were generated uniformly over $t_j \in \{0(2)40\}$, $(n = 21)$. The errors ε_j were generated from $N(0, \sigma^2)$ with mean zero and $\sigma^2 = 0.25$.

also present the estimation results using the nonlinear least squares estimator. Accordingly, we took the parameter estimates $\theta_1 = 0.4577 \times 10^{-5}$ and $\theta_2 = 0.2797 \times 10^{-3}$ from Bellman et al. (1967) together with the initial condition $\xi = x(0) = 0$, thus $\eta_0 = (\xi, \theta_1, \theta_2)^T$. Then we generated observations uniformly over $t_j \in \{0(2)40\}$, $(n = 21)$, according to (2), where the i.i.d. measurement errors ε_j were generated from the normal distribution $N(0, \sigma^2)$ with mean zero and variance $\sigma^2 = 0.25$. The solution to (16) with these parameter values is plotted in Figure 3 with a solid line, while the observations are indicated by pluses. This setup was chosen to mimic the real data scenario related to this model, as described later on.

The fact that θ_1 and θ_2 are small numbers, combined with the fact that their magnitudes are rather different, renders their estimation a difficult task, cf. p. 1303 in Esposito and Floudas (2000). In Table 6 we see the empirical average of point estimates and the empirical coverage of interval estimates based on Monte Carlo study consisting of 500 runs. The point estimates are given by (12), while the interval estimates are defined in (14). We note that when estimating $\theta = (\theta_1, \theta_2)$, unlike Bellman et al. (1967), Van Domselaar and Hemker (1975), Tjoa and Biegler (1991) and Varah (1982), we did not assume that the initial condition $x(0) = 0$ was known, but estimated it as well. Notice also that our method exploits linearity in the parameters and therefore it is not required to provide it with an initial guess in the parameter space (in Bellman et al. (1967) and other related papers the initial guesses $\theta_1 = 10^{-6}$ and $\theta_2 = 10^{-4}$ were used). We see that even with a small sample as 21 observations, the point and interval estimates are satisfying, and again, we do not observe a substantial difference between the ACCEL and NLS methods.

We next tested our approach on the real data for the model (16) given in Table 39 in Bodenstein (1922) and reproduced in Table I in Bellman et al. (1967). There are in total 14 observations available on the interval $[0, 39]$, excluding the initial condition $x(0) = 0$. [2] This time we did not estimate the initial condition and considered it to be

[2]Note that in Table 39 in Bodenstein (1922) and in Table I in Bellman et al. (1967) the observation 48.8 corresponding to the time instance $t = 19$ appears to contain a typo: we tentatively corrected it to 38.8. The

Table 6: Means of point estimates and actual coverage of interval estimates for the parameters of model (16), where the initial value ξ is zero and the parameters are $\theta_1 = 0.4577 \times 10^{-5}$ and $\theta_2 = 0.2797 \times 10^{-3}$. The results are based on 500 simulation runs. There are 21 observations given on a uniform grid on $[0,40]$, and the errors are normal with zero expectation and $\sigma^2 = 0.25$. The ACCEL point estimates are given by (12); the interval estimates are defined in (14).

			ACCEL		NLS	
Setup			Mean	Coverage	Mean	Coverage
n=21	ξ_1	0	1.491e-02	0.938	7.960e-03	0.942
	θ_1	4.577e-06	4.576e-06	0.954	4.577e-06	0.952
	θ_2	2.797e-04	2.788e-04	0.932	2.798e-04	0.930

zero, which agrees with the physical phenomenon the model describes. The estimation results are displayed in Table 7. Both point and interval estimates obtained through the ACCEL and NLS methods are presented. The results are very similar.

Table 7: Point estimates for the parameters of model (16) based on the real data of Table 39 in Bodenstein (1922). We consider the initial value to be zero. The ACCEL point estimates are given by (12); the confidence intervals were generated according to (14). The left and right interval points are denoted by CI(L) and CI(R), respectively.

		ACCEL			NLS	
	Point	CI(L)	CI(R)	Point	CI(L)	CI(R)
θ_1	4.579e-06	4.255e-06	4.903e-06	4.577e-06	4.253e-06	4.901e-06
θ_2	2.791e-04	1.923e-04	3.658e-04	2.796e-04	1.928e-04	3.665e-04

A comparison to the results given in Table 5 shows that this is essentially the same result as already reported in the literature using the least squares estimator: this illustrates the fact that just one iteration in the Newton-Raphson type procedure suffices to arrive at an asymptotically equivalent estimator to the least squares estimator of the parameter of interest, provided a preliminary estimator is already within the $n^{-1/2}$ range of the true parameter. In Figure 4 we plot the data from Bellman et al. (1967) and the solution to (16) evaluated with ACCEL fitted values of θ_1 and θ_2. The fit appears to be satisfactory given a simplistic character of the model (16).

4.2 Barnes' problem

This problem was studied in Varah (1982), who refers to Van Domselaar and Hemker (1975) for exact experimental details. The system under study represents chemical reaction equations and is actually a variant of the Lotka-Volterra system that was treated

same correction was applied in Table 24 in Esposito and Floudas (2000) and in Table 1 in Kim and Sheng (2010).

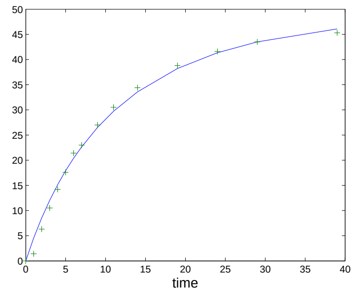

Figure 4: The solution to (16) is given by the solid line, while the observations are indicated by pluses. The parameters were estimated using the real data from Bellman et al. (1967). The initial value is considered to be known and equals zero.

above. The system takes the form

$$\begin{cases} x_1'(t) = \theta_1 x_1(t) - \theta_2 x_1(t) x_2(t), \\ x_2'(t) = \theta_2 x_1(t) x_2(t) - \theta_3 x_2(t). \end{cases} \tag{17}$$

In the real data of Varah (1982) there are 11 data points which is clearly a rather small sample. Actually, the sample size maybe too small to draw significant conclusions, but nevertheless, the example is still interesting. Table 8 presents simulation results, trying to mimic the real data scenario. We generate 11 equidistant observations on the interval $[0,5]$ according to (17) and (2), where the errors are normal with mean zero and standard deviation $\sigma = 0.05$. The NLS method seems to perform somewhat better in this specific scenario, although the difference with ACCEL is not dramatic.

Table 8: Means of point estimates and actual coverage of interval estimates for the parameters of model (17), where the initial value ξ is considered as known. The results are based on 500 simulation runs. There are 11 observations given on a uniform grid on $[0,5]$, and the errors are normal with zero expectation and $\sigma = 0.05$. The ACCEL point estimates are given by (12); the interval estimates are defined in (14).

			ACCEL		NLS	
Setup			Mean	Coverage	Mean	Coverage
n=11	θ_1	8.600e-01	8.652e-01	0.898	8.600e-01	0.944
	θ_2	2.079e+00	2.082e+00	0.924	2.078e+00	0.948
	θ_3	1.624e+00	1.626e+00	0.930	1.623e+00	0.946

In Table 9 we see the resulting point and interval estimates based on the real data, using the ACCEL and NLS methods. The solutions of the system (17) corresponding to these parameter estimates are displayed in Figure 5.

As Van Domselaar and Hemker (1975), we consider the initial values to be known, and use the first observation of each system state for that. We note that Van Domselaar and Hemker

Table 9: Point estimates for the parameters of model (17) based on the real data appears in Varah (1982). We consider the initial values to be known. The ACCEL point estimates are given by (12); the confidence intervals were generated according to (14). The left and right interval points are denoted by CI(L) and CI(R), respectively.

	ACCEL			NLS		
	Point	CI(L)	CI(R)	Point	CI(L)	CI(R)
θ_1	5.139e-01	1.607e-01	8.672e-01	8.609e-01	7.508e-01	9.709e-01
θ_2	1.552e+00	7.877e-01	2.317e+00	2.079e+00	1.898e+00	2.260e+00
θ_3	1.316e+00	6.066e-01	2.026e+00	1.815e+00	1.624e+00	2.006e+00

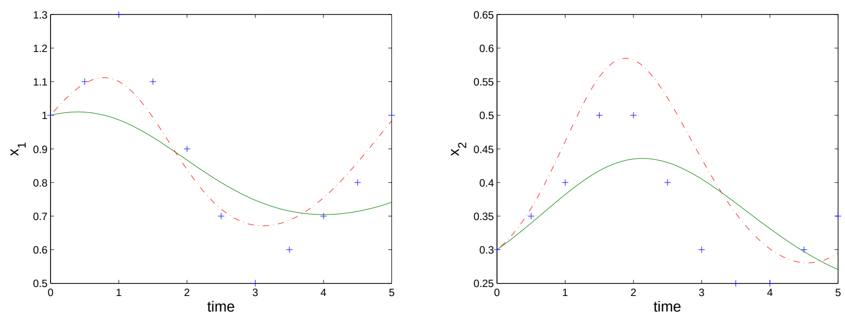

Figure 5: The solution to (17) based on the ACCEL (solid line) and the NLS (dash dotted line) estimates; the observations are indicated by pluses. The parameters were estimated using the real data from Varah (1982).

(1975) used an initial guess in the parameter space, while using the direct integral approach as we do here does not requires such a guess. Varah (1982) points out that the data are only accurate to about 10%, and thus one should not try too hard to fit the data closely. Specifically, he obtained different parameter estimates, depending on the knots position he used for the spline approximation of the trajectories. The results of the NLS here are similar to those Varah (1982) report as 'VDH values' and those reported in Van Domselaar and Hemker (1975). We note that the estimates obtained using the ACCEL method are different from those obtained using the NLS; also, the ACCEL confidence intervals are wider than the NLS ones. This is unlike all the examples we studied above where the ACCEL and NLS methods yielded almost identical results. Thus, at least in this specific example, the ACCEL method seems to be more conservative. However, as mentioned above, the sample size is likely to be too small to make any meaningful conclusions.

4.3 α-pinene problem

We now consider 'Problem 8' of Tjoa and Biegler (1991). The system is given by

$$
\begin{aligned}
x_1'(t) &= -(\theta_1+\theta_2)x_1(t),\\
x_2'(t) &= \theta_1 x_1(t),\\
x_3'(t) &= \theta_2 x_1(t) - (\theta_3+\theta_4)x_3(t) + \theta_5 x_5(t),\\
x_4'(t) &= \theta_3 x_3(t),\\
x_5'(t) &= \theta_4 x_3(t) - \theta_5 x_5(t).
\end{aligned} \tag{18}
$$

This system characterizes a reaction that describes the thermal isomerization of α-pinene (x_1) to dipentene (x_2) and alloocimene (x_3), which in turn yields $\alpha-$ and $\beta-$pyronene (x_4) and a dimer (x_5). The data we use are taken from Table 2 in Box et al. (1973). For each state of the system, the data includes only 8 observations in time. Thus, this is not a trivial problem to deal with, a point raised also in Tjoa and Biegler (1991). In Table 10 we see the resulting point and interval estimates based on the real data, using the ACCEL method. We do not present the NLS method since it did not converge in a reasonable amount of time (using straightforward optimization as we did in all examples above). In the last column we present the estimation result from Tjoa and Biegler (1991). The solution of the system (18) corresponding to the ACCEL estimate is displayed in Figure 6. Unlike Tjoa and Biegler (1991), our approach does not require to provide an initial guess in the parameter space.

Table 10: Point estimates for the parameters of model (18) based on the real data appears in Box et al. (1973). We consider the initial values to be known. The ACCEL point estimates are given by (12); the confidence intervals were generated according to (14). The left and right interval points are denoted by CI(L) and CI(R), respectively.

	Point	CI(L)	CI(R)	Tjoa and Biegler (1991)
θ_1	5.869e-05	5.771e-05	5.967e-05	5.926e-05
θ_2	2.830e-05	2.740e-05	2.920e-05	2.963e-05
θ_3	1.745e-05	1.305e-05	2.186e-05	2.047e-05
θ_4	2.132e-04	1.770e-04	2.494e-04	2.744e-04
θ_5	2.137e-05	1.037e-05	3.236e-05	3.997e-05

We conducted two simulation studies, corresponding to two different measurement error variances. Specifically, we generate observations according to (2) and (18) with the following setup: the time grid is the same as in the real data, namely $t_j \in \{1230, 3060, 4920, 7800, 10680, 15030, 22620, 36420\}$, resulting in a total of 8 observation points. Initial values are set to the observations at the first time point, $\xi = \{88.35, 7.3, 2.3, 0.4, 1.75\}$. The errors are normal with expectation zero and standard deviations $\sigma = a \times \{44.6833, 36.4111, 4.9570, 1.6339, 12.4147\}$ corresponding to σ_i, $i = 1, \dots, 5$. Here, the value a is multiplied by the mean value of each state, as calculated from the solutions based on the real data example. In the first study we set $a = 0.02$, while in the second we take $a = 0.1$. We note that the variance σ^2 that corresponds to $a = 0.02$

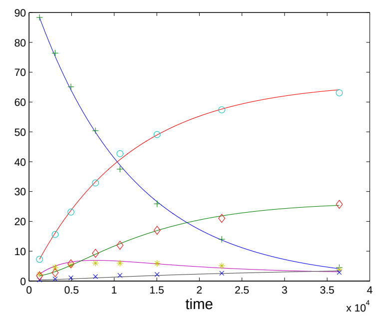

Figure 6: The solution to (18) based on the ACCEL estimates; the observations are indicated by different symbols, corresponding to the system state they represent. The parameters were estimated using the real data from Box et al. (1973).

is the order of the variance that we observed in the real data example. For each scenario, we repeat the experimental setup 500 times and calculate the average of point estimates and actual coverage of the confidence intervals. We also provide the standard error of the ACCEL estimator as calculated based on 500 simulations (STE), as well as the square root of the average of estimates of the asymptotic variance (ASYM). The results are presented in Table 11. We see that the actual coverage is not too poor, but nevertheless deviates noticeably from the nominal level of 95%. Further, we see a considerable difference between estimates of the asymptotic variance and the actual finite sample variance as calculated based on 500 simulations. All these results are not surprising, if we recall that we have at hand only 8 observations for each system state.

5 Summary

Parameter estimation for ordinary differential equations is a challenging problem. Existing methods (certainly those that are frequently applied in practice) mainly revolve around the least squares estimation theme. Although in theory the least squares estimator possesses good (asymptotic) properties, its evaluation is far from trivial in the current setting. Alternatively, smoothing-based methods have been shown to escape some of the computational challenges associated with the least squares estimation, but are typically less efficient statistically, meaning informally that they do not squeeze maximal possible information from the observations. In this paper we have demonstrated on a number of real and simulated data examples that execution of a computationally simple one-step Newton-Raphson type procedure on a preliminary smoothing-based estimator leads to rather satisfactory estimation results, that are comparable to those in the least squares estimation. We remark that while it is well known that the choice of the smoothing parameter in nonparametric estimation is a rather delicate issue, here we use a very simple bandwidth selection procedure, but nevertheless the obtained results are good, so that our method does not require much fine tuning from the user. Furthermore, our estimation method enables the researcher to obtain confidence intervals. In

Table 11: Means of point estimates and actual coverage of interval estimates for the parameters of model (18), where the initial value ξ is considered as known. The results are based on 500 simulation runs; see the experimental setup in the text. The ACCEL point estimates are given by (12); the interval estimates are defined in (14). Standard errors of the point estimates as calculated based on the 500 simulations (denoted by 'STE'). Square root of the average of the estimates of the asymptotic variance, over the 500 simulations (denoted by 'ASYM').

Setup		True	Mean	Coverage	STE	ASYM
$\sigma = 0.02$	θ_1	5.926e-05	5.920e-05	0.758	6.539e-07	3.913e-07
	θ_2	2.963e-05	2.958e-05	0.806	5.246e-07	3.615e-07
	θ_3	2.047e-05	2.042e-05	1.000	5.789e-07	1.815e-06
	θ_4	2.744e-04	2.709e-04	1.000	7.847e-06	2.099e-05
	θ_5	3.997e-05	3.878e-05	0.998	2.793e-06	6.060e-06
$\sigma = 0.1$	θ_1	5.926e-05	5.910e-05	0.768	3.265e-06	3.026e-02
	θ_2	2.963e-05	2.945e-05	0.820	2.669e-06	2.717e-03
	θ_3	2.047e-05	1.993e-05	0.998	2.755e-06	9.746e-06
	θ_4	2.744e-04	2.452e-04	0.946	8.382e-05	1.406e-04
	θ_5	3.997e-05	3.103e-05	0.940	2.569e-05	9.688e-05

particular, the empirical coverage of the confidence intervals we provide is good even for samples as small as $n = 21$ in the examples we considered. On the other hand, the approach discussed in this work requires that all the components of the solution to (1) are observed. Thus in future work we will apply it to a smooth and match method that does not require a fully observed system.

All computations were carried out using Matlab. The code will be sent by the first author upon request. The local polynomial estimator in some of the examples was based on the implementation from Cao (2008).

Acknowledgements

The research leading to these results has received funding from the European Research Council under ERC Grant Agreement 320637.

References

Bellman, R., J. Jacquez, R. Kalaba, and S. Schwimmer (1967). Quasilinearization and the estimation of chemical rate constants from raw kinetic data. *Mathematical Biosciences 1*(1), 71–76.

Bellman, R. and R. S. Roth (1971). The use of splines with unknown end points in the identification of systems. *Journal of Mathematical Analysis and Applications 34*(1), 26–33.

Bodenstein, M. (1922). Bildung und zersetzung der höheren stickoxyde. *Z. Phys. Chem. 100*, :68–123.

Box, G., W. Hunter, J. MacGregor, and J. Erjavec (1973). Some problems associated with the analysis of multiresponse data. *Technometrics 15*(1), 33–51.

Brunel, N. J. B. (2008). Parameter estimation of ode's via nonparametric estimators. *Electronic Journal of Statistics 2*, 1242–1267.

Burden, R. L. and J. D. Faires (2001). Numerical analysis. *Brooks/Cole 935*.

Cao, Y. (2008). Local linear kernel regression. *MATLAB Central File Exchange (http://www.mathworks.com/matlabcentral/ fileexchange/19564-local-linear-kernel-regression), Retrieved March 18, 2015*.

Dattner, I. and C. A. J. Klaassen (2015). Optimal rate of direct estimators in systems of ordinary differential equations linear in functions of the parameters. *arXiv:1305.4126v2*.

Edelstein-Keshet, L. (2005). *Mathematical models in biology. Classics in Applied Mathematics*, Volume 46. Society for Industrial and Applied Mathematics.

Esposito, W. R. and C. A. Floudas (2000). Global optimization for the parameter estimation of differential-algebraic systems. *Industrial & Engineering Chemistry Research 39*(5), 1291–1310.

Feinberg, M. (1979). Lectures on chemical reaction networks. *Notes of lectures given at the Mathematics Research Center, University of Wisconsin*.

Floudas, C. A. (1999). *Handbook of test problems in local and global optimization*.

Gelman, A., F. Bois, and J. Jiang (1996). Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *Journal of the American Statistical Association 91*(436), 1400–1412.

Girolami, M. (2008). Bayesian inference for differential equations. *Theoretical Computer Science 408*(1), 4–16.

Girolami, M. and B. Calderhead (2011). Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73*(2), 123–214.

Gugushvili, S. and C. A. J. Klaassen (2012). $\sqrt{n}$-consistent parameter estimation for systems of ordinary differential equations: bypassing numerical integration via smoothing. *Bernoulli 18*, 1061–1098.

Himmelblau, D., C. Jones, and K. Bischoff (1967). Determination of rate constants for complex kinetics models. *Industrial & Engineering Chemistry Fundamentals 6*(4), 539–543.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, 633–643.

Kim, T. and Y. Sheng (2010). Estimation of water quality model parameters. *KSCE Journal of Civil Engineering 14*(3), 421–437.

Ramsay, J. O., G. Hooker, D. Campbell, and J. Cao (2007). Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69*(5), 741–796.

Sontag, E. D. (2001). Structure and stability of certain chemical networks and applications to the kinetic proofreading model of t-cell receptor signal transduction. *Automatic Control, IEEE Transactions on 46*(7), 1028–1047.

Swartz, J. and H. Bremermann (1975). Discussion of parameter estimation in biological modelling: algorithms for estimation and evaluation of the estimates. *Journal of Mathematical Biology 1*(3), 241–257.

Tjoa, I. B. and L. T. Biegler (1991). Simultaneous solution and optimization strategies for parameter estimation of differential-algebraic equation systems. *Industrial & Engineering Chemistry Research 30*(2), 376–385.

Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*. Springer.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

Van Domselaar, B. and P. W. Hemker (1975). Nonlinear parameter estimation in initial value problems. Technical report, SIS-76-1121.

Varah, J. (1982). A spline least squares method for numerical parameter estimation in differential equations. *SIAM Journal on Scientific and Statistical Computing 3*(1), 28–46.

Voit, E. O. (2000). *Computational analysis of biochemical systems: a practical guide for biochemists and molecular biologists*. Cambridge University Press.

Voit, E. O. and J. Almeida (2004). Decoupling dynamical systems for pathway identification from metabolic profiles. *Bioinformatics 20*(11), 1670–1681.

Vujačić, I., I. Dattner, J. González, and E. C. Wit (2014). Time-course window estimator for ordinary differential equations linear in the parameters. *Statistics and Computing, to appear*.

Wu, C.-F. (1981). Asymptotic theory of nonlinear least squares estimation. *The Annals of Statistics*, 501–513.